

## ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Special Issue: *Genetic Novelty and Genomic Variation by RNA Networks and Viruses*

PERSPECTIVE

# What viruses tell us about evolution and immunity: beyond Darwin?

Felix Broecker<sup>1</sup> and Karin Moelling<sup>2,3</sup><sup>1</sup>Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, New York. <sup>2</sup>Institute of Medical Microbiology, University of Zurich, Zurich, Switzerland. <sup>3</sup>Max Planck Institute for Molecular Genetics, Berlin, Germany

Addresses for correspondence: Felix Broecker, Department of Microbiology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029–6574. felixbroecker@gmx.net; Karin Moelling, Max Planck Institute of Molecular Genetics, Virology, Ihnestrasse 73, Berlin D-14195, Germany. moelling@molgen.mpg.de

We describe mechanisms of genetic innovation mediated by viruses and related elements that, during evolution, caused major genetic changes beyond what was anticipated by Charles Darwin. Viruses and related elements introduced genetic information and have shaped the genomes and immune systems of all cellular life forms. None of these mechanisms contradict Darwin's theory of evolution but extend it by means of sequence information that has recently become available. Not only do small increments of genetic information contribute to evolution, but also do major events such as infection by viruses or bacteria, which can supply new genetic information to a host by horizontal gene transfer. Thereby, viruses and virus-like elements act as major drivers of evolution.

**Keywords:** Darwin; viruses; evolution; horizontal gene transfer

## Beyond Darwin?

Charles Darwin revolutionized our understanding of the origin and evolution of species. The central tenets of his theory of evolution state that hereditary variation occurs slowly due to the gradual accumulation of random small modifications that are subject to natural selection. The discovery of DNA as a carrier of the hereditary information then allowed linking inherited traits to mutations in DNA. Today, we also know that RNA, with its high plasticity due to the relatively high infidelity of RNA polymerases, can exert strong evolutionary influences. New technologies for high-throughput genome sequencing have recently led to the discovery of multiple novel forms of mutations and genetic alterations.

Fundamentally new insights into these influences have come from viruses and microorganisms, infectious agents that can supply genetic material into recipient cells by horizontal gene transfer (HGT). For instance, up to two-thirds of the human genome sequence is derived from viruses and transposable elements,<sup>1,2</sup> and a smaller proportion of DNA sequence originates from bacteria and

other microorganisms.<sup>3–5</sup> In contrast to point mutations, small insertions, or small deletions, which have conventionally been regarded as the major driving forces of evolution, the supply of new genes by HGT can cause dramatic changes to an organism almost instantaneously. In addition to HGT, major genetic changes are induced by crossover and recombination events, (retro)transposition activity, transformation, and conjugation;<sup>3–6</sup> well-studied examples include uptake of plasmids that confer antibiotic resistance to a bacterial cell.<sup>7</sup> Furthermore, RNA agents, including RNA viruses and viroids, typically lack proofreading mechanisms during their replication and thus can accumulate mutations more rapidly than replicating DNAs.<sup>8,9</sup> Retroviruses, with their high genomic plasticity and life cycle that includes obligatory integration into the host genome, are among the major drivers of evolution by, among other things, providing novel genes and regulatory elements.<sup>10–13</sup> Retroviruses have accumulated in large numbers into eukaryotic genomes, mediate HGT, transduce cellular genes including oncogenes, and contribute to cancer.<sup>1,2,10</sup>

doi: 10.1111/nyas.14097

Ann. N.Y. Acad. Sci. 1447 (2019) 53–68 © 2019 The Authors. *Annals of the New York Academy of Sciences* published by Wiley Periodicals, Inc. on behalf of New York Academy of Sciences.

53

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

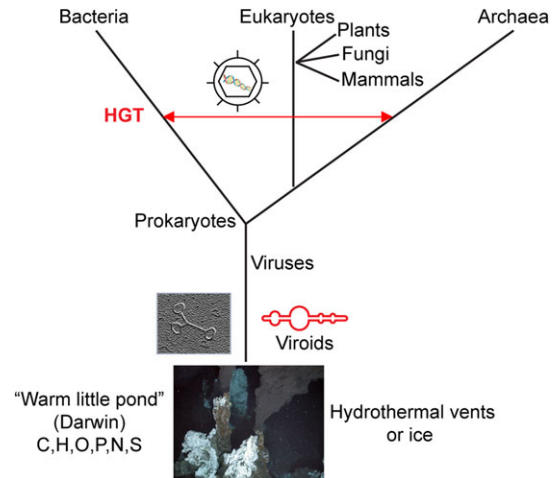
Genetic transfer events between different species across all domains of the tree of life, and cross-talk between various genomes (cellular and viral), raise new questions on the definition of a species.<sup>14</sup> The conventional definition of a species assumes that any organism acquires its set of genes solely from its parents (subject to minor variations in each generation due to imperfect copying mechanisms). However, mechanisms like HGT can introduce genetic material even from distantly related species, including viruses or smaller genetic agents such as transposable elements.<sup>14</sup>

Microbial and viral infections, furthermore, can lead to endosymbiotic relationships and are at the origin of eukaryotic structures such as mitochondria and chloroplasts (both originating from formerly free-living bacteria), and possibly the nucleus, which may have originated from a complex virus.<sup>15–17</sup> Viral infections were crucial for the evolution of the mammalian placenta<sup>11</sup> and various immune systems in prokaryotes and eukaryotes,<sup>18–20</sup> and they shaped genomes throughout evolution—genomes that now serve as archives of previous viral infections.<sup>21</sup>

Single-stranded RNA (ssRNA) is the most innovative biomolecule known. We can learn about its properties by analyzing ribozymes (catalytic RNA species) and closely related virus-like viroids. An entire living world could be built from RNA only, whereby RNA could replicate, cleave, join, form peptide bonds, and evolve depending on environmental changes.<sup>22,23</sup> Evolutionary innovations were then fixed in the form of a more stable and less variable molecule, double-stranded DNA (dsDNA), which likely arose after RNA. The known DNA polymerases typically introduce fewer mutations than do RNA polymerases, due to efficient proof-reading mechanisms.<sup>8,9</sup> DNA, however, can also change significantly by mechanisms such as mobility of transposable elements (DNA transposons and retrotransposons). Transposable elements can be regarded as capsid-free viruses<sup>24</sup> that are trapped within host cells but occasionally can be transferred between different cellular organisms.<sup>25,26</sup>

## The evolutionary tree of life

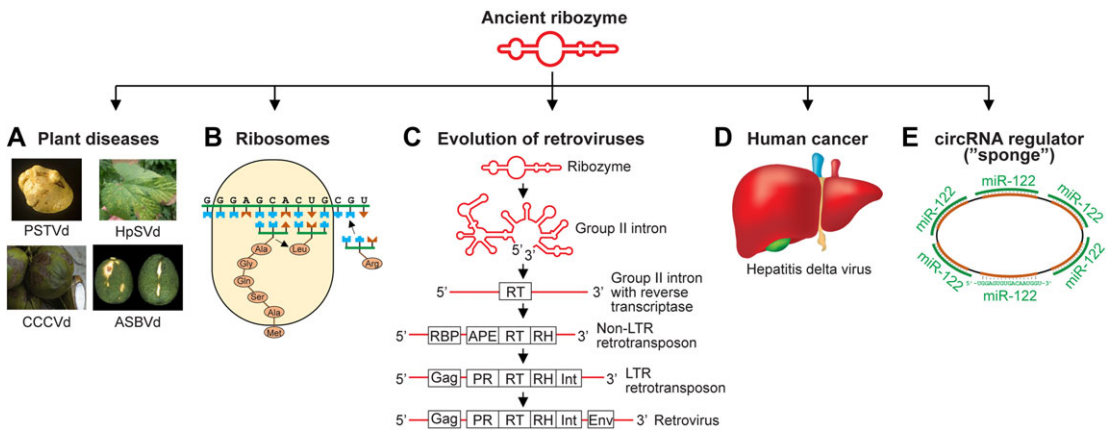
A simplified tree of life is shown in Figure 1. The three arms represent phylogenetic relationships between the bacterial, archaeal, and eukaryotic domains, as inferred by ribosomal DNA



**Figure 1.** Phylogenetic tree based on ribosomal DNA sequences<sup>27</sup> (not to scale). The red line indicates horizontal gene transfer (HGT) by viruses and related elements.

sequences.<sup>27</sup> The horizontal line symbolizes HGT events among the three arms. HGT has been a frequent phenomenon during evolution and is often mediated by viruses and other selfish genetic elements.<sup>28–31</sup> For example, the human genome contains up to two-thirds of sequences originating from retroviruses and transposable elements, and at least 145 genes (of a total of 21,000) are likely derived from bacteria, archaea, fungi, protists, and plants.<sup>1,2,5</sup> Viruses infect organisms of all three domains of cellular life “from root to the leaves,” but due to their polyphyletic nature are often not included in the tree of life.<sup>32</sup> We, however, hypothesize viruses to be at the root.<sup>33</sup> This is speculative but based on the fact that virus-like elements, such as viroids, are the smallest known replicating entities with structural and not coding information. Catalytically-active and replicating species can arise from populations of RNA with random sequences or quasispecies.<sup>23</sup> Furthermore, the recently discovered giant viruses indicate that the evolutionary transition from virus to cell may be a continuum;<sup>34</sup> and ribosomes have likely evolved from simple ribozymes in the ancient RNA world.<sup>35</sup> Thus, although there are opposing views,<sup>36</sup> we<sup>21,33</sup> and others<sup>37,38</sup> think that viruses must be located far down at the root of the tree of life.

Bacteria, archaea, and eukaryotes are normally described as originating from the last universal common ancestor, which requires DNA, protein



**Figure 2.** Simple ribozymes evolved into various structures. (A) Viroids are related to ribozymes, some have retained ribozyme activity. PSTVd, Potato spindle tuber viroid; HpSVd, Hop stunt viroid; CCCVd, Coconut cadang-cadang viroid; ASBVd, Avocado sunblotch viroid. (B) A ribozyme became the catalytically active center of ribosomes.<sup>47</sup> (C) Ribozymes may have evolved gradually to today’s retroviruses by acquiring coding information.<sup>50</sup> APE, apurinic endonuclease; Gag, group-specific protein; Env, envelope; Int, integrase; PR, protease; RBP, RNA-binding protein; RH, RNase H; RT, reverse transcriptase. (D) The hepatitis delta virus likely originated from a plant viroid that picked up coding information from a cell.<sup>51</sup> (E) Circular RNAs (circRNAs) are involved in gene regulation by binding to short regulatory RNAs comparable to a “sponge” and have been identified in all domains of life, suggesting an ancient evolutionary origin.<sup>54,55</sup>

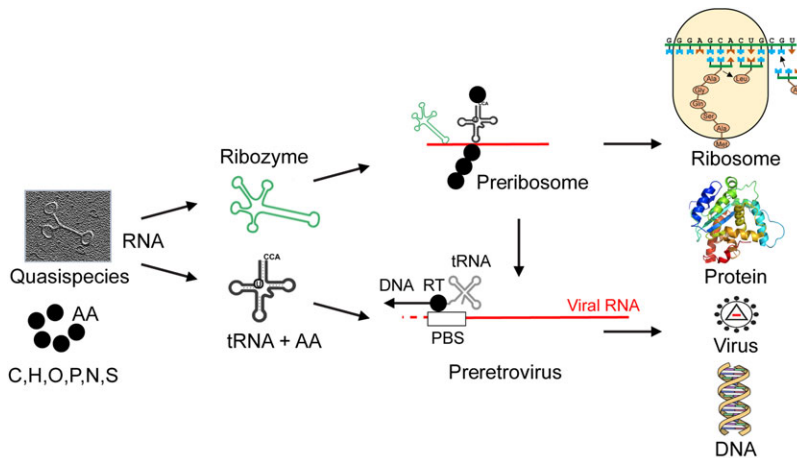
synthesis, and must, therefore, have evolved for long periods from simpler forms of life. The origin of life on Earth, in general, is speculative but most likely included forms of noncoding replicating RNA-like molecules as the first biopolymers, possibly linked with phosphodiester bridges or more stable peptide linkages as peptide nucleic acid.<sup>39,40</sup> In the test tube, catalytically active RNA molecules have been selected from a pool of ~10<sup>15</sup> RNA polymers 220 nt in length with random sequences that can adopt a multitude of conformations, including ribozymes and tRNA-like structures described as quasispecies.<sup>23,41,42</sup> Ribozymes do not encode proteins but contain structural information through hairpin–loop folds, can perform a multitude of catalytic reactions, including RNA cleavage and ligation, as well as peptide bond formation; they can mutate and thereby adjust to environmental changes and replicate without the help of protein-based polymerases.<sup>22,23</sup>

Ribozymes are closely related to viroids, circular naked RNA virus-like elements that can cause diseases in plants, such as the Potato spindle tuber viroid (Fig. 2A).<sup>43,44</sup> Some plant viroids found today have retained ribozyme activity, but many have lost it; they are replicated by RNA polymerases of the host cell. Loss of function is an often-underestimated evolutionary force facilitated by the

supply of functions through the environment, causing a reduction of genome size to reduce the energy required for replication.

The first ribozymes may have formed at hydrothermal vents at the bottom of the oceans in niches with large surface areas and concentrating effects where minerals supply metal ions as catalysts and high temperatures accelerate chemical reactions.<sup>45</sup> However, even ice can support ribozyme activity, as the progressive dehydration during ice crystal formation concentrates the substrates (ribonucleotide triphosphates and magnesium salts), while ribozyme activity is retained even below 0 °C.<sup>46</sup> Later, during cellular evolution, ribozymes became integral parts of ribosomes and constitute the enzymatically active component for protein synthesis by forming peptide bonds: “Ribosomes are ribozymes” (Fig. 2B).<sup>47</sup> There are examples of ribozymes that acquired coding information during evolution. For example, it has been suggested that a reverse transcriptase gene became inserted into group II introns,<sup>48,49</sup> which then may have further evolved into retroviruses<sup>50</sup> (Fig. 2C).

A rare case of a circular RNA (circRNA) with coding capacity is the hepatitis delta virus (HDV), which presumably originated from a noncoding plant viroid. HDV integrated a gene, likely of cellular origin, into its catalytic ribozyme (Fig. 2D).<sup>51</sup>



**Figure 3.** Hypothetical models for the evolution from ribozyme to ribosomes and retroviruses.

HDV contributes to liver cancer in combination with hepatitis B virus, which incorporates HDV into its virions to allow for transmission to other cells and organisms.<sup>51,52</sup> Another viroid-like RNA with protein-coding information is the smallest known naturally occurring replicating RNA only 220 nt in length, the small circular satellite of rice yellow mottle virus (scRYMV) that is made up completely of coding sequences.<sup>53</sup> The origin of the protein sequences is unclear. scRYMV can be considered as an evolutionary intermediate from simple ribozymes toward coding RNAs.

Recently, circRNAs have been described as molecules involved in mammalian gene regulation by binding to a number of short regulatory RNAs comparable to an absorbing “sponge” (Fig. 2E).<sup>54</sup> circRNAs have subsequently been identified in all domains of cellular life,<sup>55</sup> suggesting an evolutionary ancient origin. Alternatively, modern circRNAs may have several independent origins, yet all are based on noncoding RNA (ncRNA).

tRNA-like folds may have been among the earliest RNA structures together with ribozymes and have played a central role in the evolution of replication and protein synthesis (Fig. 3).<sup>56,57</sup> Originally, tRNA-like structures likely evolved as a tag to initiate RNA genome replication in the RNA world before the advent of proteins.<sup>56</sup> This functionality can still be observed in some contemporary viruses where tRNAs and tRNA-like structures are essential for replication.<sup>58</sup> For example, a number

of plant positive-strand RNA viruses bear tRNA-like structures carrying one specific amino acid at the 3'-termini, which—among other functions—act as primers to initiate minus-strand RNA synthesis by the viral replicase.<sup>58</sup> Cellular tRNAs also act as primers for the reverse transcriptase of retroviruses, plant pararetroviruses, and LTR retrotransposons, to initiate DNA synthesis.<sup>59</sup>

tRNAs may have also played a role in the formation of early retrovirus-like structures, selecting for a tRNA-binding site on evolving RNAs. This combination is maintained in extant retroviruses where tRNAs serve as primers to initiate replication *via* a DNA copy. In addition, precursor tRNAs likely evolved into ribosomal RNAs which bear many striking similarities to tRNA sequences.<sup>50,60</sup> In present-day cellular organisms, tRNAs serve as the carriers of amino acids to the ribosome to enable protein synthesis. Highlighting the versatility of the tRNA fold, other functions have also been described. For example, tRNA-derived small RNAs have been shown to regulate gene expression and suppress transposable elements in eukaryotes.<sup>61,62</sup> In addition, tRNAs are considered as a relevant component in transgenerational inheritance through germline cells (see below).

It may be interesting to note that not only ribozymes but also deoxyribozymes likely evolved before the emergence of a reverse transcriptase, through chemical removal of oxygen from ribonucleotides.<sup>23</sup>

## Viruses as drivers of evolution

Influenza viruses, whose negative-sense ssRNA genomes are segmented into eight parts that are copackaged into the virion, provide a good example for two types of evolutionary processes. On the one hand, the surface glycoproteins, subject to immune pressure by the host, undergo incremental point mutations (as proposed by Darwin), a process termed antigenic drift.<sup>63</sup> On the other hand, entire gene segments can be exchanged if a cell is coinfecting with two or more different influenza viruses, a process called antigenic shift. Antigenic drift causes subtle antigenic alterations that help the virus to gradually evade host immunity. Antigenic shift, on the other hand, can instantaneously cause a drastic antigenic change. The introduced glycoproteins may be entirely unknown to the immune systems of large parts of the population. Therefore, antigenic shift events (and not antigenic drift) were at the origin of the major influenza virus pandemics such as the 2009 H1N1 “swine flu,” a triple reassortant virus containing genome segments originating from human, swine, and avian viruses.<sup>64</sup> Influenza viruses were among the first to indicate the difference between point mutations and gene exchange.<sup>65</sup> Antigenic shift thereby is a form of HGT between related influenza viruses.

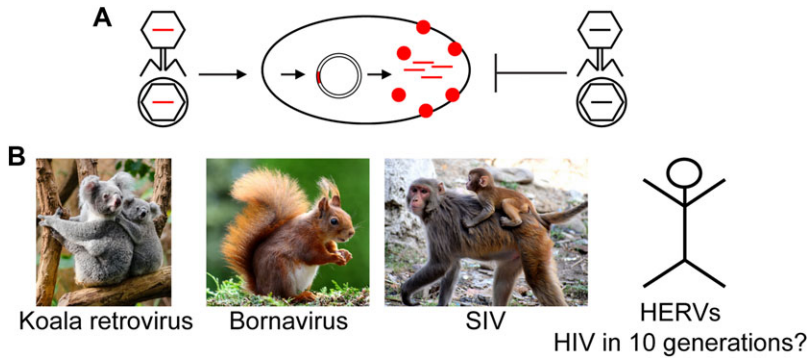
Retroviruses can also undergo subtle mutational changes that are facilitated by the infidelity of their reverse transcriptase.<sup>10,65</sup> This helps the virus to efficiently escape the host immunity. More drastic genetic changes can result from recombination events with cellular genes, including oncogenes, as described below. Each retroviral infection, following reverse transcription of the ssRNA genome into a dsDNA provirus, supplies the host cell's genome with a set of new genes, *gag*, *pol*, and *env* and auxiliary genes, as in the case of lentiviruses like HIV.<sup>10,65</sup> Retroviral insertion can induce downstream activation of cellular genes by the retroviral promoter, or insertional inactivation.<sup>13,66,67</sup> Such events are a potential cause of chronic diseases, including cancer. For example, the expression of the *c-Myc* proto-oncogene can be dysregulated through the nearby insertion of avian leukosis virus, leading to B cell lymphomas in chickens.<sup>68</sup> Infections of somatic cells only affect one individual since the integrated virus is not passed on to the next generation. Some retroviruses, however, can integrate into germline

cells and are consequently inherited by future generations. These are the endogenous retroviruses that populate eukaryotic genomes.<sup>65</sup> Endogenous retroviruses can accumulate to high numbers in genomes through reinfection and/or intracellular amplification.<sup>1,10,12</sup> In rare cases, other RNA viruses, such as Bornaviruses, and DNA viruses can be integrated (see below and Fig. 4). Subsequently, viral genes can be captured by the host cell to exert novel functions, as described below for the syncytins.

## Oncogenes and cancer

Retroviruses are potent mediators of HGT, which was initially discovered by oncogenes that are taken up by recombination between the integrated viral genome and cellular genes.<sup>65</sup> Many oncogenes were selected for by their strong growth promoting ability *in vitro* and a high potential for malignant transformation or tumor growth in animal models.<sup>10,69,70</sup> Some of the products of oncogenes, such as ErbB, Ras, Raf, or Myc, are dysregulated in human cancers and are targets for anticancer therapy.<sup>70</sup> The infidelity of retroviral replication by the error-prone reverse transcriptase often leads to mutations, such that cellular and viral oncogenes can differ. For example, the *v-Src* oncogene of Rous sarcoma virus (RSV) is a C-terminally truncated variant of the cellular *c-Src* proto-oncogene. The truncation eliminates a regulatory domain such that the *v-Src* protein causes malignant transformation, whereas *c-Src* with the intact regulatory domain does not.<sup>71</sup> Thus, RSV serves as a vehicle for HGT, carrying a mutated variant of the *c-Src* gene to other cells.

Oncogenes are often taken up by retroviruses at the expense of replicating genes, which leads to non-replicating defective viral entities that require the supply of the lacking genes of other superinfecting retroviruses *in trans* to infect other cells.<sup>10,21,69</sup> The generally modular structure of viral genomes reflects how they have evolved by gene uptake from various sources.<sup>72</sup> This allows for *de novo* recombination of genes occurring naturally or artificially by genetic engineering, which is the basis for the design of viral gene therapy vectors carrying therapeutic genes, or tailor-made phages for antibacterial purposes.<sup>73</sup>



**Figure 4.** Endogenization of viruses causes immunity. (A) Schematic representation of superinfection exclusion. A phage/virus (genomes indicated as red lines) infects a cell and the genome is integrated into the cellular genome. RNA or proteins from the endogenized phage/virus protect against superinfections by related invaders (genomes shown as black lines). (B) Examples for superinfection exclusion include koala retrovirus,<sup>74–76</sup> bornaviruses in ground squirrels,<sup>94</sup> and simian immunodeficiency virus (SIV).<sup>77</sup> In humans, human endogenous retroviruses (HERVs) may cause resistance against related retroviruses including HIV.<sup>153</sup> In addition, it has been suggested that endogenization of HIV itself can happen<sup>80</sup> and may cause resistance to AIDS in the future.

## Endogenous viruses

The process of endogenization has been studied recently in koalas where an ongoing endogenization of a retrovirus into the germline is observed in real time (Fig. 4).<sup>74–76</sup> In the early 1900s, some koalas were isolated on islands off the Australian coast as they were endangered to go extinct. The animals on mainland Australia (and some of the islands) were infected via a cross-species transmission with a virus related to gibbon ape leukemia virus.<sup>74–76</sup> Many animals succumbed to the virus termed koala retrovirus (KoRV). However, due to its capacity to infect germline cells, KoRV became endogenized in some koalas.<sup>74–76</sup> The endogenized viruses likely protect the animals against *de novo* infections with exogenous KoRV.<sup>74,75</sup> Since koala populations on some islands remained free of KoRV, it is likely that the first endogenization event has occurred within about the last 100 years.<sup>74</sup> Endogenization is still ongoing as a significant degree of interindividual variability is observed with respect to numbers and integration sites of KoRV elements.<sup>74</sup> KoRV is an example for an integrated retrovirus that may protect against *de novo* infection by similar viruses, a mechanism designated as superinfection exclusion that has been described for other retroviruses, and also nonretroviruses, as well.<sup>18,50</sup> Superinfection exclusion can be based, for instance, on destruction of the invading viral genes by RNase H-like nucleases or the expression of viral gene products that block cell surface receptors.

Protection by a virus against superinfecting viruses is presumably also the case in some monkey species that allow for replication of simian immunodeficiency virus (SIV), a close relative of HIV, but do not succumb to the infection.<sup>77</sup> These monkeys are presumably resistant because their immune systems were shaped by lentiviruses such that they became resistant to SIV-induced immunodeficiency.<sup>77</sup> One can speculate that this may also happen with HIV in humans in the future, which will likely take at least 10 generations, as judged from the koalas, whose generation time is about 7.8 years. HIV has indeed been shown to be able to infect germline cells, which would be a requirement for endogenization into the germline.<sup>78,79</sup> In addition, there is evidence that defective HIV proviruses protect human T cells from superinfection with HIV.<sup>80</sup>

During evolution, retroviral endogenization happened frequently enough to populate the human genome with 450,000 retroviral elements that have typically accumulated inactivating mutations if the expression of viral genes was not selected for. The human genome consists of almost 50% of full-length or truncated endogenous retroviruses and related retroelements.<sup>1</sup> Some human genes, like protein kinase inhibitor beta (*PKIB*), consist of up to 80% of such elements, usually located in introns.<sup>81</sup> This raises the question about what the upper limit of retrovirus-related insertions may be in eukaryotic genomes. The majority of endogenous

retroviruses have been degenerated by homologous recombination of the long terminal repeats (LTRs) and elimination of the internal sequences.<sup>1,13</sup> These solitary LTRs are so abundant that the total genetic information would be longer than the human genome itself if they were full-length retroviruses of about 10 kilobases.

As the proof that the ancient retroviral insertions in the human genome are derived from formerly infectious retroviruses, a consensus sequence from the sequences of nine full-length endogenized retroviruses was constructed.<sup>82</sup> The sequence was converted into a synthetic virus, designated as Phoenix, which was able to form virus particles and infect mammalian cells *in vitro*.<sup>82</sup>

A major benefit for human ancestors was germline infections with two retroviruses now designated as human endogenous retrovirus (HERV) types W and FRD.<sup>83,84</sup> The corresponding infectious retroviruses entered the ancestral germline about 30 and 45 million years ago, respectively. Gene products of these viruses, syncytins, which are derived from the retroviral envelope protein and related to HIV gp41, are involved in placental development.<sup>11</sup> Specifically, the coordinated expression of syncytins 1 and 2 is required for the formation of the syncytiotrophoblast that establishes the interface between maternal blood and embryonic extracellular fluid. In addition, syncytins mediate the protection of the fetus from maternal immune rejection through the immunosuppressive domain.<sup>85</sup> Similarly, HIV has the ability to fuse cells via its gp41 protein, which leads to syncytia formation by cell fusion. In the case of syncytins, ancient retroviruses induce immune suppression for the benefit of the host.

Other than providing novel genes to the host, endogenous retroviruses and other retroelements have been shown to regulate host gene expression by providing promoters, enhancers, transcription factor binding sites, and other regulatory elements. We have shown that some LTRs of the HERV-K(HML-10) family of HERVs, which invaded the ancestral genome about 35 million years ago, have retained promoter activity.<sup>13</sup> Moreover, one of the LTRs primes a regulatory transcript that suppresses the expression of the death-associated protein-3 gene (*DAP3*) implicated in apoptosis.<sup>13</sup> Knock-down of the retroviral transcript was sufficient to induce apoptosis in cancer cell lines *in vitro*,

suggesting that the LTR may contribute to the apoptosis-evading phenotype of some cancer cells. However, a single endogenous retrovirus is unlikely a major cause of cancer; rather it may contribute to the phenotype of specific cancerous cells. Similar regulatory transcripts originating from other LTRs in the human genome likely contribute to the complex network of human gene regulation and possibly to cancer.<sup>13</sup> In addition, endogenous retroviruses have been shown to provide regulatory elements including transcription factor binding sites to mammalian genes, which, for example, have been shown to modulate innate immunity pathways.<sup>12</sup>

Recently, HERV-K has been suspected to contribute to various human diseases, such as cancer and amyotrophic lateral sclerosis,<sup>86</sup> and reverse transcriptase activity, likely of retroelement origin, to Alzheimer's disease (AD)<sup>87</sup> (see below).

In addition to endogenous retroviruses, mammalian genomes harbor sequences derived from *Circoviridae*, *Filoviridae*, *Bornaviridae*, *Parvoviridae*, *Herpesviridae*, and others.<sup>88–92</sup> A well-studied example of the function of endogenous non-retroviral sequences are Borna disease viruses (BDVs), negative-sense ssRNA viruses that cause fatal encephalitis in sheep, horses, and cattle.<sup>88,93</sup> Horses, for instance, are susceptible to Bornaviral disease, which includes signs of depression. Intriguingly, these susceptible species have no detectable BDV-related sequences in their genomes.<sup>88</sup> In contrast, less susceptible species, including primates, mice, and rats, have genomic BDV sequences.<sup>88</sup> Indeed, experimental evidence for a protective role of endogenous BDV sequences against BDV infection has been obtained in squirrels.<sup>94</sup> Various other virus-derived sequences in the genomes of mammals and other species have been recruited for antiviral and immune-related functions, as described elsewhere.<sup>12,18–20</sup>

Noncoding RNA in the human genome amounts to about 98% of the sequence, leaving only about 2% for protein-coding information.<sup>1</sup> In contrast, it is low or close to zero in viruses or prokaryotes.<sup>95</sup> In RNA viruses, the 5'-termini are often noncoding with structural, not genetic information, such as dimerization signals in retroviral genomes, the transactivation response element in HIV or internal ribosomal entry sites found in many positive-sense ssRNA viruses.<sup>65</sup>

## Transposable elements

Related to viruses are transposable elements found in pro- and eukaryotic genomes.<sup>24</sup> The transposable elements are either DNA transposons exerting a cut-and-paste transposition or retroelements, which undergo reverse transcription and integration at a new genomic locus by a copy-and-paste mechanism that leads to gene duplication—a major property of complex genomes.<sup>1,96</sup> Gene duplication is a major innovative step since one copy can maintain its functions, while the other one can change. Transpositions can cause major changes in the host's genome, and may compromise its integrity, as integrations are mutagenic events.<sup>67</sup> Highly abundant retrotransposons are indicators of active evolution and gain of functions of an organism which has to cope with new environmental conditions. This seems to be the case in marine ecosystems, as retrotransposon activity of ocean plankton is apparently high, with reverse transcriptase genes amounting to up to 13.5% of the metagenomes.<sup>97</sup> Also, RNase H-like genes are of high frequency in marine plankton, with about 15 copies per cell on average.<sup>50</sup> Evolutionary events mediated by transposable elements appear to be constantly ongoing in the oceans and remain to be studied in detail.<sup>97</sup>

Another measure for the activity of transposable elements is the expression level of transposases.<sup>98</sup> Transposases can introduce innovation in bacterial genomes and thereby promote adaptation to dynamic environments. They not only influence the host cell's genome, but can also mediate HGT between cells.<sup>99</sup> In marine bacteria, transposase expression levels are particularly high under environmental stress such as low oxygen as well as in biofilms and particles.<sup>98</sup> This may relate to higher rates of genome modifications and HGT necessary to adapt to environmental stress, whereby dense populations may provide hotspots of bacterial evolution.<sup>98</sup>

In eukaryotes, transposon activity is counteracted by RNA interference (RNAi).<sup>100</sup> In stem cells and sperm cells of various species, such as mice or *Caenorhabditis elegans*, and likely also in humans, transposon activity is silenced by a dedicated RNAi mechanism, the Piwi-interacting RNA (piRNA) pathway.<sup>101</sup> Piwi proteins mediate the destruction of RNA from transposable elements, guided by piRNAs, 26–31 nt in length. They are part of the

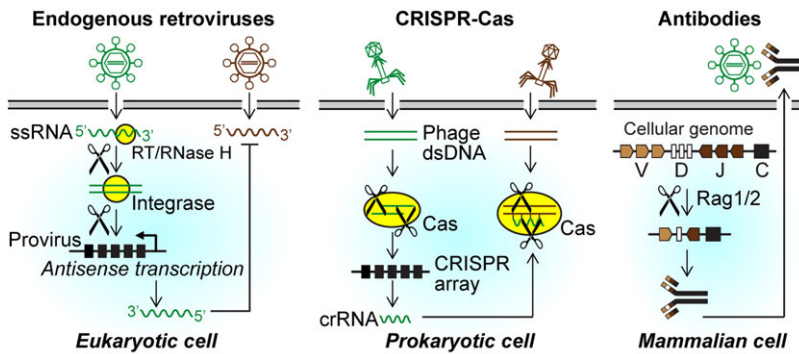
Argonaute protein family, which have the same three-dimensional structure as the retroviral RNase H.<sup>102</sup> This silencing mechanism suppresses transposon activity in stem and germline cells, where too high transposon activity may result in the accumulation of detrimental genetic defects; too much “jumping” would result in an “error catastrophe,” as defined by Manfred Eigen, leading to cell death.<sup>41,103,104</sup> Inactivation of Piwi activity can lead to male infertility, highlighting the importance of transposable element silencing.<sup>105</sup> The frequency of piRNAs as mediators of transgenerational inheritance transmitted *via* sperm cells is presently a matter of debate.<sup>106</sup>

## Related immune mechanisms in eukaryotes and prokaryotes

Cellular organisms have evolved a plethora of antiviral and antiphage defense systems.<sup>107,108</sup> It is worth comparing antiviral mechanisms mediated by viruses in eukaryotes (“viruses protect against viruses”)<sup>12,18–20,109</sup> with the defense systems of prokaryotes where “phages protect against phages.”<sup>18,110</sup> Many mechanisms are analogous and can be described as the endogenization of invading sequences to defend related invaders. In this regard, HERV-mediated protection and CRISPR/Cas9 are analogous.<sup>18</sup> Also, superinfection exclusion and viral interference are related defense mechanisms found in both eukaryotes and prokaryotes.<sup>18</sup> A central role in many cellular immune systems is played by RNase H-like proteins, such as the Cas9 nuclease in CRISPR/Cas9, the Piwi domain in Argonaute-based silencing, and the RAG1 recombinase that mediates V(D)J recombination to diversify antibody and T cell receptor sequences (Fig. 5).<sup>18,81,111</sup>

RNases H are typically not sequence specific, but their cleavage sites depend on fused domains, which in the case of HIV is the reverse transcriptase, together with structural properties of the polypurine tract, a preferred binding site in the RNA template.<sup>112</sup> This specificity creates the dinucleotide overhangs essential for subsequent dsDNA provirus integration by the integrase, which also exhibits an RNase H-fold.<sup>113</sup> The Cas9 enzyme of CRISPR/Cas9 is an RNase H-related enzyme directed by a guide RNA to cleave the DNA strand in RNA–DNA hybrids, as in this case defense is directed against an invading phage DNA.<sup>114</sup> RNases





**Figure 5.** Endogenization of retroviruses and phages into the host genomes to serve as antiviral defense is orthologous in pro- and eukaryotes (crRNA stands for CRISPR RNA). The generation of diversity of antibodies by V(D)J recombination also involves virus-like elements, *Transib* transposons.<sup>111</sup>

H can even cleave an RNA strand in dsRNA, highlighting the versatility of this protein fold.<sup>115</sup>

In addition, the components of the host antiviral defense systems, in prokaryotes or eukaryotes, such as RNAi-type defense, are similar to the set of gene products that make up retroviruses.<sup>81,109</sup> These similarities are, among others, the lengths of the cleavage products of around 20 nt, staggered cleavage events with dinucleotide overhangs, the common fold of Piwi and RNase H domains, nucleic acid binding by Argonaute and reverse transcriptase, unwinding and helicase activities, Dicer and integrase, and caspase and protease; all of these are orthologous, as described.<sup>81,109</sup>

Antiviral responses can be based not only on cleavage enzymes guided by nucleic acids, but also on blocking infection through receptor occupation by viral gene products as well as transcriptional or translational inhibition.<sup>18</sup> Viruses and genetic parasites as inventors of antiviral systems have also been described by Villarreal.<sup>19,20</sup>

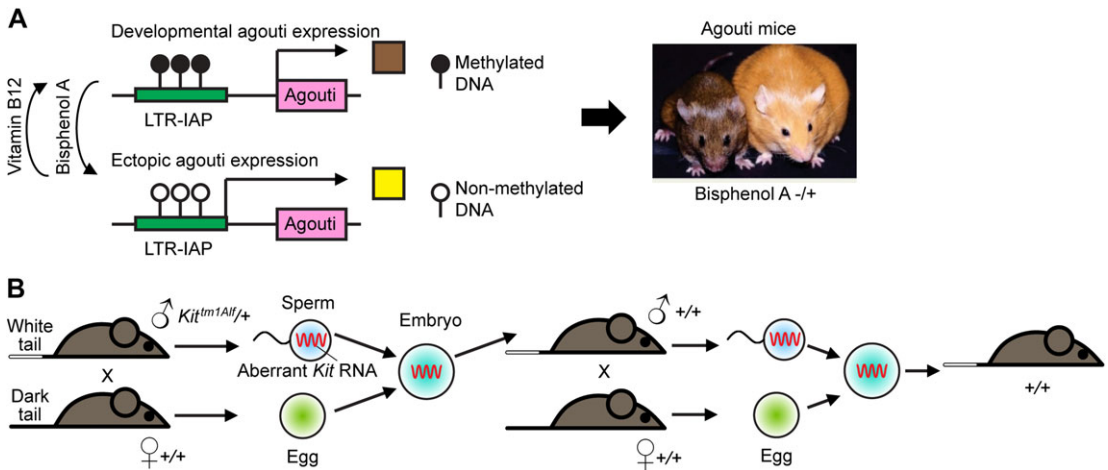
The reverse transcriptase, first discovered in retroviruses, is highly abundant in almost all cellular species. There are more than a thousand bacterial variants with mostly unknown functions and reverse transcriptases are also abundantly expressed in marine plankton.<sup>97,116</sup> Reverse transcriptase activity in the human brain may play a role in the development of AD, affecting the amyloid precursor protein gene (*APP*).<sup>87</sup> Recombination of the gene with different reverse transcribed “genomic cDNAs” of various full-length or truncated spliced mRNAs was more frequently found in the brains of AD patients than in healthy controls.<sup>87</sup>

The genomic cDNAs resemble retrotransposition events and pseudogene formation by long interspersed nuclear elements (LINEs) that encode a reverse transcriptase.<sup>87</sup> Indeed, LINE-1 retrotransposition events have been detected in the human brain, creating a mosaicism of somatic *de novo* LINE-1 insertions.<sup>117</sup> It has been suggested that this mosaicism may have evolved to increase the complexity and diversity of the mammalian brain.<sup>118</sup>

The reverse transcriptase can also be part of diversity-generating retroelements (DGRs) that are found in bacterial and phage genomes.<sup>119</sup> DGRs utilize the infidelity of the reverse transcriptase to diversify DNA sequences. The prototype DGR of *Bordetella* bacteriophage BPP-1 is able to modify the phage’s receptor specificity and thereby alter tropism.<sup>119</sup>

## Epigenetics and paramutations

One of the mechanisms of mammalian organisms to control the activity of endogenous retroviruses is epigenetic silencing of the LTRs.<sup>120</sup> In some cases, such epigenetic modifications have been recruited to regulate host genes. These can cause major changes in phenotypes beyond Darwin. A well-studied animal model for epigenetic effects is the inbred agouti mouse in which *agouti* expression depends on the degree of methylation of an alternative promoter located in the LTR of an intracisternal A particle (IAP), a type of endogenous retrovirus.<sup>121</sup> Hypomethylation of the LTR leads to strong *agouti* expression, causing mice to have yellow fur and to develop obesity (Fig. 6). Hypermethylation prevents *agouti* expression, resulting in brown,



**Figure 6.** Epigenetics and paramutations as examples of non-Mendelian inheritance. (A) Environmental effects on fur color of *agouti* mice by vitamin B12 or bisphenol A on methylation of the alternative long terminal repeat (LTR) promoter of an intracisternal A particle (IAP), epigenetic modifications that are inheritable.<sup>121,122</sup> (B) Transgenerational regulation of expression of *Kit* by RNA-mediated inheritance *via* sperm cells.<sup>132</sup> Heterozygous male mice with a *Kit<sup>tm1Alf</sup>* allele (a null allele that makes no functional Kit protein but an aberrant *Kit* RNA; a “paramutagenic” allele) and a normal wild-type (+) allele have a tail with a white tip. The aberrant *Kit* RNA is packaged into sperm cells and transmitted to the embryo. Progeny carry a wild-type *Kit* allele transmitted by the father, but action of the transmitted aberrant *Kit* RNA still gives rise to the white tail. This allows for paramutation of the wild-type *Kit* allele and further transmission of the white tail. Loss of aberrant RNA through dilution over successive generations might lead to gradual loss of the paramutation.

nonobese mice. The diet influences methylation of the LTR and thereby the mouse phenotype. Obese yellow mice can become lean brown mice by supplementing their diet with methyl donors, such as vitamin B12 or folic acid.<sup>122</sup> In contrast, the toxin bisphenol A causes LTR hypomethylation; exposed mice become yellow and obese.<sup>121</sup> Another gene susceptible to the level of methylation of an IAP is the *axin fused* allele that influences tail kinking in mice.<sup>123</sup> The total number of IAP elements in the mouse genome is about 1000. About 100 of them have been identified recently as being variably methylated in C57BL/6 mice, they might, therefore, be involved in epigenetic regulation of gene expression, similar to *agouti* and *axin fused*.<sup>124,125</sup> In germline cells, epigenetic marks are normally erased and remodeled in a process termed *epigenetic reprogramming*.<sup>120</sup> Therefore, epigenetic changes are mostly transient and only affect one individual in its lifetime. However, in some cases, modifications may escape eradication and they can be passed on to the offspring, whereby the altered epigenetic modifications typically return to the normal state within a few generations by dilution.<sup>126</sup> For example, the epigenetic marks regulating *agouti*,

*axin fused*, and a third locus have been shown to be transgenerationally inheritable.<sup>121,123,124</sup> Epigenetic inheritance not directly mediated by DNA modifications but by small RNAs, termed paramutations, is presumably more stable over time.<sup>127</sup> Small RNA-mediated paramutations (see below) can last for at least 20 generations in the nematode *C. elegans*.<sup>128</sup> Recently, tRNAs and R-loop formation have also been found as mechanisms that contribute to epigenetic changes of gene expression.<sup>129,130</sup>

The phenomenon of paramutations was discovered in maize by the Canadian plant geneticist Royal Alexander Brink in the 1960s.<sup>131</sup> Paramutations do not follow Mendelian laws and are, in contrast to transient changes, nongenetic modifications that are inherited and long-lasting over multiple generations.<sup>132</sup> Paramutations have recently been attributed in some cases to various RNAs species, including piRNAs, miRNAs, tRNA-derived small RNAs, long ncRNAs, and others, which are transmitted through the sperm of mice.<sup>132–137</sup> Transgenerational inheritance was shown, for instance, to have an impact on the tail phenotype of mice and was identified as being mediated by miRNAs specific for the

tyrosine kinase proto-oncogene *Kit*, transmitted through the sperm.<sup>132</sup> Similarly, in mice, cardiac hypertrophy,<sup>133</sup> body size,<sup>134</sup> obesity,<sup>135</sup> metabolic syndrome,<sup>127</sup> and stress-induced gene expression programs<sup>136,137</sup> were found to be influenced by transgenerational inheritance mediated by miRNAs, with the phenotypic effects usually lasting over several generations. It has recently been suggested, however, that transgenerational inheritance in mammals may be a relatively rare phenomenon.<sup>106</sup>

## Endosymbiosis

Endosymbiosis, or symbiogenesis, is another phenomenon where major genetic innovation can occur through a single event. Symbiogenesis is defined by the introduction of new genetic information by uptake of an entire organism, whereby the symbiotic relationship is extended beyond a “master and slave” interaction.<sup>138</sup> Symbiotic relationships between partners with bacteria and viruses have been important events during evolution,<sup>138</sup> including flowering plants.<sup>139,140</sup> The uptake of bacteria leading to the evolution of chloroplasts and mitochondria, for example, are considered unique events during eukaryotic evolution. The engulfed bacteria lost or delegated the majority of genes, around 3000, to the host cell nucleus such that the human mitochondria, for instance, only retained 37 genes that are specialized to energy production.<sup>141,142</sup> Mitochondria likely originated from *Rickettsiales* bacteria<sup>143</sup> and chloroplasts are related to cyanobacteria.<sup>144</sup> These events, however, are not as rare as thought and similar uptakes of bacteria into other cells may still happen today. For example, bacteria of the *Rickettsia* genus lost the ability for autonomous replication and became dependent on the host.<sup>145</sup> Thus, genomes do not only increase in size during evolution, but there are also counteracting forces, loss of genes as exemplified by endosymbionts and *Rickettsia*. This may be an energy-saving measure of symbionts compared to autonomously living species. The endosymbiotic *Wolbachia* spp. bacteria can even transfer genes into their arthropod hosts and influence host reproduction and speciation.<sup>146</sup> In addition, the nucleus, a hallmark of eukaryotic cells, may originate from a giant virus related to poxviruses or mimiviruses.<sup>15–17</sup> Giant viruses exclusively replicate in the cytoplasm and may have been autonomous entities originally,<sup>24</sup> similar to the precursors of

mitochondria and other plastids. Further supporting evidence for viral nucleogenesis comes from phenotypic similarities of viral factories (the site of replication of giant viruses in cellular cytoplasm) with the nucleus.<sup>17</sup> The nucleus may have emerged by fusion of the cellular genome with the giant virus genome, which then became the nucleus, losing the ability for virus production.<sup>17</sup>

The mechanism of endosymbiosis indicates that replicating entities can give up an autonomous lifestyle and become parasitic within a richer environment, which is accompanied by the reduction of genetic information.<sup>147</sup> We, therefore, speculate that viruses may have originally been autonomous agents that later became intracellular parasites, which, however, cannot be proven.<sup>33</sup>

It appears that symbiogenesis of viruses is not a rare phenomenon either. Viral endosymbionts have been described, for instance, in insects such as *Drosophila melanogaster*.<sup>148</sup> Insects are frequently infected with sigma viruses and *D. melanogaster* has evolved several mechanisms of resistance against the endosymbiotic viruses, yet the putative benefits to the host remain unclear.<sup>148</sup> Interestingly, sigma viruses are likely purely vertically transmitted, indicating an adaptation to an endosymbiotic lifestyle and the loss of the ability of horizontal transmission.<sup>148</sup> One example of how endosymbiotic viruses can be beneficial to their host, parasitoid wasps that lay eggs into other insects, is polydnviruses.<sup>139</sup> The viruses carry unusual genomes with characteristics of eukaryotic genes and protect the wasp's larvae from immune reactions by their insect host.<sup>139</sup>

Viral endosymbionts may also influence reproduction as in the case of an RNA virus found in the moth *Homona magnanima* that appears to be associated with late male killing.<sup>149</sup> It has been suggested that viral endosymbionts may be an important but largely overlooked force in host evolution that require further investigation.<sup>148</sup>

## Conclusions

One can speculate that life on Earth started with the smallest known autonomously self-replicating catalytically active entities, the ribozymes, which consist of noncoding circRNAs with information solely based on structure. Especially, the ribozymes suggest that ribozymes preceded prokaryotes because they supply the protein synthesis

machinery in all forms of cellular life.<sup>47</sup> Possibly, viruses, or more specifically viroid-like agents, may have been among the first biological entities. Viruses may have started as simple, autonomously replicating elements, which after the emergence of cell-like containments at the beginning of life lost genes to become obligate intracellular parasites. This kind of gene reduction is a frequent but often ignored evolutionary mechanism.<sup>147</sup> Gene reduction was demonstrated *in vitro* by “Spiegelman’s Monster,” where a long coding phage RNA in the presence of an RNA polymerase, free nucleotides, and salts degenerated to a small extremely fast replicating ncRNA after 74 generations, demonstrating that loss of genes for the benefit of fast replication can occur.<sup>150,151</sup> *Rickettsia*, mitochondria, and chloroplasts are other examples for loss of genes resulting in an exclusively intracellular lifestyle.<sup>141,142,145</sup>

Possibly, viruses gave up their independence for a parasitic lifestyle. In this case, viruses could be our oldest ancestors.<sup>21,33</sup> In this article, we sought to provide possible explanations as to how early ribozymes/viroids or virus-like entities may have abandoned a former autonomy for a parasitic lifestyle, as observed today. Viruses are nowadays the most abundant biological entities on Earth, amounting to about  $10^{33}$  particles.<sup>152</sup> Non-coding RNAs, many of viral origin, have become major regulators of the coding genes in cellular organisms, which only constitute about 2% of mammalian genomes, whereas ncRNA represents about 98%.<sup>1,95</sup> The circular noncoding RNAs, circRNAs, are sponge-like absorbing agents that regulate gene expression—they structurally resemble and might be evolutionarily related to the evolutionarily ancient ribozymes and viroids.<sup>54,55</sup> The importance of viruses and parasitic elements during genome evolution and antiviral responses has been pointed out by Villarreal before.<sup>19,20,37</sup>

There are some informative intermediates combining ncRNA and coding RNA, such as the group II introns that have taken up a reverse transcriptase gene.<sup>116</sup> This is an important step of evolution, the transition from ncRNA to partially coding sequences—which may have led to the evolution of retroviruses.<sup>50</sup>

Recently, small ncRNA was identified as a genetic factor transmitted via sperm cells as a transgenerational carrier.<sup>132</sup> Future studies will have to assess

how frequent this phenomenon is. The similarities between viruses and phages as discussed here with respect to antiviral defense, shown for HERVs and CRISPR, demonstrate the close relatedness of the prokaryotic and the eukaryotic worlds.<sup>18</sup> These similarities have often been overlooked. We may be able to compare the two seemingly distant worlds and find more similarities with respect to viruses, antiviral defense, viral counter-defense, and others.

## Acknowledgments

We would like to thank Professor Peter Palese of the Icahn School of Medicine at Mount Sinai, New York, for his generous support. F.B. thanks the German Academy of Sciences Leopoldina for a postdoctoral stipend.

## Competing interests

The authors declare no competing interests.

## References

- Lander, E.S., L.M. Linton, B. Birren, *et al.* 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- de Koning, A.P., W. Gu, T.A. Castoe, *et al.* 2011. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **7**: e1002384.
- Riley, D.R., K.B. Sieber, K.M. Robinson, *et al.* 2013. Bacteria-human somatic cell lateral gene transfer is enriched in cancer samples. *PLoS Comput. Biol.* **9**: e1003107.
- Salzberg, S.L., O. White, J. Peterson, *et al.* 2001. Microbial genes in the human genome: lateral transfer or gene loss? *Science* **292**: 1903–1906.
- Crisp, A., C. Boschetti, M. Perry, *et al.* 2015. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol.* **16**: 50.
- Thomas, C.M. & K.M. Nielsen. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **3**: 711–721.
- San Millan, A. 2018. Evolution of plasmid-mediated antibiotic resistance in the clinical context. *Trends Microbiol.* **26**: 978–985.
- Drake, J.W., B. Charlesworth, D. Charlesworth & J.F. Crow. 1998. Rates of spontaneous mutation. *Genetics* **148**: 1667–1686.
- Rajon, E. & J. Masel. 2011. Evolution of molecular error rates and the consequences for evolvability. *Proc. Natl. Acad. Sci. USA* **108**: 1082–1087.
- Coffin, J.M., S.H. Hughes & H.E. Varmus, Eds. 1997. *Retroviruses*. Cold Spring Harbor Laboratory Press.
- Lavialle, C., G. Cornelis, A. Dupressoir, *et al.* 2013. Paleovirology of ‘syncytins’, retroviral env genes exapted for a role in placentation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**: 21120507.

12. Chuong, E.B., N.C. Elde & C. Feschotte. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**: 1083–1087.
13. Broecker, F., R. Horton, J. Heinrich, *et al.* 2016. The intron-enriched HERV-K(HML-10) family suppresses apoptosis, an indicator of malignant transformation. *Mob. DNA* **7**: 25.
14. Bapteste, E., E. Susko, J. Leigh, *et al.* 2005. Do orthologous gene phylogenies really support tree-thinking? *BMC Evol. Biol.* **5**: 33.
15. Bell, P.J. 2001. Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *J. Mol. Evol.* **53**: 251–256.
16. Takemura, M. 2001. Poxviruses and the origin of the eukaryotic nucleus. *J. Mol. Evol.* **52**: 419–425.
17. Forterre, P. & M. Gaïa. 2016. Giant viruses and the origin of modern eukaryotes. *Curr. Opin. Microbiol.* **31**: 44–49.
18. Broecker, F. & K. Moelling. 2019. Evolution of immune systems from viruses and transposable elements. *Front. Microbiol.* **10**: 51.
19. Villarreal, L.P. 2009. The source of self: genetic parasites and the origin of adaptive immunity. *Ann. N.Y. Acad. Sci.* **1178**: 194–232.
20. Villarreal, L.P. 2011. Viral ancestors of antiviral systems. *Viruses* **3**: 1933–1958.
21. Moelling, K. 2013. What contemporary viruses tell us about evolution: a personal view. *Arch. Virol.* **158**: 1833–1848.
22. Lincoln, T.A. & G.F. Joyce. 2009. Self-sustained replication of an RNA enzyme. *Science* **323**: 1229–1232.
23. Wilson, D.S. & J.W. Szostak. 1999. *In vitro* selection of functional nucleic acids. *Annu. Rev. Biochem.* **68**: 611–647.
24. Koonin, E.V. & V.V. Dolja. 2014. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiol. Mol. Biol. Rev.* **78**: 278–303.
25. Peccoud, J., V. Loiseau, R. Cordaux, *et al.* 2017. Massive horizontal transfer of transposable elements in insects. *Proc. Natl. Acad. Sci. USA* **114**: 4721–4726.
26. Ivancevic, A.M., R.D. Kortschak, T. Bertozzi, *et al.* 2018. Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. *Genome Biol.* **19**: 85.
27. Woese, C.R., O. Kandler & M.L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* **87**: 4576–4579.
28. Keeler, P.J. & J.D. Palmer. 2008. Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* **9**: 605–618.
29. Schaack, S., C. Gilbert & C. Feschotte. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol.* **25**: 537–546.
30. McDaniel, L.D., E. Young, J. Delaney, *et al.* 2010. High frequency of horizontal gene transfer in the oceans. *Science* **330**: 50.
31. Rosenberg, E. & I. Zilber-Rosenberg. 2018. The hologenome concept of evolution after 10 years. *Microbiome* **6**: 78.
32. Forterre, P. 2015. The universal tree of life: an update. *Front. Microbiol.* **6**: 717.
33. Moelling, K. 2012. Are viruses our oldest ancestors? *EMBO Rep.* **13**: 10333.
34. Claverie, J.M. & C. Abergel. 2013. Open questions about giant viruses. *Adv. Virus Res.* **85**: 25–56.
35. Fox, G.E. 2010. Origin and evolution of the ribosome. *Cold Spring Harb. Perspect. Biol.* **2**: a003483.
36. Moreira, D. & P. López-García. 2009. Ten reasons to exclude viruses from the tree of life. *Nat. Rev. Microbiol.* **7**: 306–311.
37. Villarreal, L.P. & G. Witzany. 2010. Viruses are essential agents within the roots and stem of the tree of life. *J. Theor. Biol.* **262**: 698–710.
38. Koonin, E.V., T.G. Senkevich & V.V. Dolja. 2006. The ancient Virus World and evolution of cells. *Biol. Direct* **1**: 29.
39. Joyce, G.F. 2002. The antiquity of RNA-based evolution. *Nature* **418**: 214–221.
40. Zhang, B. & T.R. Cech. 1997. Peptide bond formation by *in vitro* selected ribozymes. *Nature* **390**: 96–100.
41. Eigen, M. 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**: 465–523.
42. Brackett, D.M. & T. Dieckmann. 2006. Aptamer to ribozyme: the intrinsic catalytic potential of a small RNA. *ChemBiochem* **7**: 839–843.
43. Diener, T.O. 1991. Viroids and viroid-like satellite RNAs: a phylogenetic analysis. *Prog. Clin. Biol. Res.* **364**: 243–256.
44. Flores, R., S. Gago-Zachert, P. Serra, *et al.* 2014. Viroids: survivors from the RNA world? *Annu. Rev. Microbiol.* **68**: 395–414.
45. Szostak, J.W. 2016. On the origin of life. *Medicina (B. Aires)* **76**: 199–203.
46. Attwater, J., A. Wochner, V.B. Pinheiro, *et al.* 2010. Ice as a protocellular medium for RNA replication. *Nat. Commun.* **1**: 76.
47. Cech, T.R. 2000. Structural biology. The ribosome is a ribozyme. *Science* **289**: 878–879.
48. Mohr, G., P.S. Perlman & A.M. Lambowitz. 1993. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Res.* **21**: 4991–4997.
49. Zimmerly, S. & C. Semper. 2015. Evolution of group II introns. *Mob. DNA* **6**: 7.
50. Moelling, K., F. Broecker, G. Russo, *et al.* 2017. RNase H as gene modifier, driver of evolution and antiviral defense. *Front. Microbiol.* **8**: 1745.
51. Taylor, J.M. 2009. Replication of the hepatitis delta virus RNA genome. *Adv. Virus Res.* **74**: 103–121.
52. Magnusius, L., J. Taylor, W.S. Mason, *et al.* 2018. ICTV virus taxonomy profile: deltavirus. *J. Gen. Virol.* **99**: 1565–1566.
53. AbouHaidar, M.G., S. Venkataraman, A. Golshani, *et al.* 2014. Novel coding, translation, and gene expression of a replicating covalently closed circular RNA of 220 nt. *Proc. Natl. Acad. Sci. USA* **111**: 14542–14547.
54. Hansen, T.B., T.I. Jensen, B.H. Clausen, *et al.* 2013. Natural RNA circles function as efficient microRNA sponges. *Nature* **495**: 384–388.

55. Danan, M., S. Schwartz, S. Edelheit, *et al.* 2012. Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.* **40**: 3131–3142.
56. Weiner, A.M. & N. Maizels. 1999. The genomic tag hypothesis: modern viruses as molecular fossils of ancient strategies for genomic replication, and clues regarding the origin of protein synthesis. *Biol. Bull.* **196**: 327–328.
57. Ariza-Mateos, A. & J. Gómez. 2017. Viral tRNA mimicry from a biocommunicative perspective. *Front. Microbiol.* **8**: 2395.
58. Dreher, T.W. 2009. Role of tRNA-like structures in controlling plant virus replication. *Virus Res.* **139**: 217–229.
59. Marquet, R., C. Isel, C. Ehresmann & B. Ehresmann. 1995. tRNAs as primer of reverse transcriptases. *Biochimie* **77**: 113–124.
60. Root-Bernstein, R., Y. Kim, A. Sanjay, *et al.* 2016. tRNA evolution from the proto-tRNA minihelix world. *Transcription* **7**: 153–163.
61. Martinez, G., S.G. Choudury & R.K. Slotkin. 2017. tRNA-derived small RNAs target transposable element transcripts. *Nucleic Acids Res.* **45**: 5142–5152.
62. Sablok, G., K. Yang, R. Chen, *et al.* 2017. tRNA derived smallRNAs: smallRNAs repertoire has yet to be decoded in plants. *Front. Plant Sci.* **8**: 1167.
63. Krammer, F., G.J.D. Smith, R.A.M. Fouchier, *et al.* 2018. Influenza. *Nat. Rev. Dis. Primers* **4**: 3.
64. Trifonov, V., H. Khiabani & R. Rabadan. 2009. Geographic dependence, surveillance, and origins of the 2009 influenza A (H1N1) virus. *N. Engl. J. Med.* **361**: 115–119.
65. Flint, S.J., V.R. Racaniello, G.F. Rall, *et al.* 2015. *Principles of Virology*. 4th ed. Washington, DC: ASM Press.
66. Hayward, W.S., B.G. Neil & S.M. Astrin. 1981. Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukemia. *Nature* **290**: 475–480.
67. Soriano, P., T. Gridley & R. Jaenisch. 1987. Retroviruses and insertional mutagenesis in mice: proviral integration at the *Mov 34* locus leads to early embryonic death. *Genes Dev.* **1**: 366–375.
68. Fan, H. & C. Johnson. 2011. Insertional oncogenesis by non-acute retroviruses: implications for gene therapy. *Viruses* **3**: 398–422.
69. Hanahan, D. & R.A. Weinberg. 2011. Hallmarks of cancer: the next generation. *Cell* **144**: 646–674.
70. Vogt, P.K. 2012. Retroviral oncogenes: a historical primer. *Nat. Rev. Cancer* **12**: 639–648.
71. Broecker, F., C. Hardt, R. Herwig, *et al.* 2016. Transcriptional signature induced by a metastasis-promoting c-Src mutant in a human breast cell line. *FEBS J.* **283**: 1669–1688.
72. Koonin, E.V., V.V. Dolja & M. Krupovic. 2015. Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology* **479–480**: 2–25.
73. Kilcher, S., P. Studer, C. Muessner, *et al.* 2018. Cross-genus rebooting of custom-made, synthetic bacteriophage genomes in L-form bacteria. *Proc. Natl. Acad. Sci. USA* **115**: 567–572.
74. Tarlinton, R.E., J. Meers & P.R. Young. 2006. Retroviral invasion of the koala genome. *Nature* **442**: 79–81.
75. Tarlinton, R., J. Meers & P. Young. 2008. Biology and evolution of the endogenous koala retrovirus. *Cell. Mol. Life Sci.* **65**: 3413–3421.
76. Stoye, J.P. 2006. Koala retrovirus: a genome invasion in real time. *Genome Biol.* **7**: 241.
77. Heeney, J.L., A.G. Dalgleish & R.A. Weiss. 2006. Origins of HIV and the evolution of resistance to AIDS. *Science* **313**: 462–466.
78. Bagasra, O., H. Farzadegan, T. Seshamma, *et al.* 1994. Detection of HIV-1 proviral DNA in sperm from HIV-1-infected men. *AIDS* **8**: 1669–1674.
79. Nuovo, G.J., J. Becker, A. Simsir, *et al.* 1994. HIV-1 nucleic acids localize to the spermatogonia and their progeny. A study by polymerase chain reaction *in situ* hybridization. *Am. J. Pathol.* **144**: 1142–1148.
80. Colson, P., I. Ravaut, C. Tamalet, *et al.* 2014. HIV infection en route to endogenization: two cases. *Clin. Microbiol. Infect.* **20**: 1280–1288.
81. Moelling, K. & F. Broecker. 2015. The reverse transcriptase-RNase H: from viruses to antiviral defense. *Ann. N.Y. Acad. Sci.* **1341**: 126–135.
82. Dewannieux, M., F. Harper, A. Richaud, *et al.* 2006. Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. *Genome Res.* **16**: 1548–1556.
83. Mi, S., X. Lee, X. Li, *et al.* 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**: 785–789.
84. Blaise, S., N. de Parseval, L. Bénil, *et al.* 2003. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc. Natl. Acad. Sci. USA* **100**: 13013–13018.
85. Mangeney, M. & T. Heidmann. 1998. Tumor cells expressing a retroviral envelope escape immune rejection *in vivo*. *Proc. Natl. Acad. Sci. USA* **95**: 14920–14925.
86. Garcia-Montojo, M., T. Doucet-O'Hare, L. Henderson, *et al.* 2018. Human endogenous retrovirus-K (HML-2): a comprehensive review. *Crit. Rev. Microbiol.* **44**: 715–738.
87. Lee, M.H., B. Siddoway, G.E. Kaeser, *et al.* 2018. Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature* **563**: 639–645.
88. Belyi, V.A., A.J. Levine & A.M. Skalka. 2010. Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. *PLoS Pathog.* **6**: e1001030.
89. Belyi, V.A., A.J. Levine & A.M. Skalka. 2010. Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the parvoviridae and circoviridae are more than 40 to 50 million years old. *J. Virol.* **84**: 12458–12462.
90. Horie, M., T. Honda, Y. Suzuki, *et al.* 2010. Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature* **463**: 84–87.
91. Katzourakis, A. & R.J. Gifford. 2010. Endogenous viral elements in animal genomes. *PLoS Genet.* **6**: e1001191.
92. Aswad, A. & A. Katzourakis. 2014. The first endogenous herpesvirus, identified in the tarsier genome, and novel

- sequences from primate rhadinoviruses and lymphocryptoviruses. *PLoS Genet.* **10**: e1004332.
93. Carbone, K.M. 2001. Borna disease virus and human disease. *Clin. Microbiol. Rev.* **14**: 513–527.
  94. Fujino, K., M. Horie, T. Honda, *et al.* 2014. Inhibition of Borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome. *Proc. Natl. Acad. Sci. USA* **111**: 13175–13180.
  95. Mattick, J.S., R.J. Taft & G.J. Faulkner. 2010. A global view of genomic information—moving beyond the gene and the master regulator. *Trends Genet.* **26**: 21–28.
  96. Cerbin, S. & N. Jiang. 2018. Duplication of host genes by transposable elements. *Curr. Opin. Genet. Dev.* **49**: 63–69.
  97. Lescot, M., B. Hingamp, K.K. Kojima, *et al.* 2016. Reverse transcriptase genes are highly abundant and transcriptionally active in marine plankton assemblages. *ISME J.* **10**: 1134–1146.
  98. Vigil-Stenman, T., K. Ininbergs, B. Bergmann, *et al.* 2017. High abundance and expression of transposases in bacteria from the Baltic Sea. *ISME J.* **11**: 2611–2623.
  99. Frost, L.S., R. Leplae, A.O. Summers, *et al.* 2005. Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* **3**: 722–732.
  100. Obbard, D.J., K.H. Gordon, A.H. Buck, *et al.* 2009. The evolution of RNAi as a defence against viruses and transposable elements. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**: 99–115.
  101. Kawano, M., H. Kawaji, V. Grandjean, *et al.* 2012. Novel small noncoding RNAs in mouse spermatozoa, zygotes and early embryos. *PLoS One* **7**: e44542.
  102. Majorek, K.A., S. Dunin-Horkawicz, K. Steczkiewicz, *et al.* 2014. The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification. *Nucleic Acids Res.* **42**: 4160–4179.
  103. Sumper, M. & R. Luce. 1975. Evidence for *de novo* production of self-replicating and environmentally adapted RNA structures by bacteriophage Qbeta replicase. *Proc. Natl. Acad. Sci. USA* **72**: 162–166.
  104. Eigen, M. 2002. Error catastrophe and antiviral strategy. *Proc. Natl. Acad. Sci. USA* **99**: 13374–13376.
  105. Vagin, V.V., A. Sigova, C. Li, *et al.* 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**: 320–324.
  106. Isbel, L. & D. Schübeler. 2018. Non-mendelian inheritance in mammals is highly constrained. *Cell* **175**: 1179–1181.
  107. tenOver, B.R. 2016. The evolution of antiviral defense systems. *Cell Host Microbe* **19**: 142–149.
  108. Rostøl, J.T. & L. Marraffini. 2019. (Ph)ighting phages: how bacteria resist their parasites. *Cell Host Microbe* **25**: 184–194.
  109. Moelling, K., A. Matskevich & J.S. Jung. 2006. Relationship between retroviral replication and RNA interference machineries. *Cold Spring Harb. Symp. Quant. Biol.* **71**: 365–368.
  110. Bondy-Denomy, J., J. Qian, E.R. Westra, *et al.* 2016. Prophages mediate defense against phage infection through diverse mechanisms. *ISME J.* **10**: 2854–2866.
  111. Kapitonov, V.V. & E.V. Koonin. 2015. Evolution of the RAG1-RAG2 locus: both proteins came from the same transposon. *Biol. Direct* **10**: 20.
  112. Charneau, P., G. Mirambeau, P. Roux, *et al.* 1994. HIV-1 reverse transcription. A termination step at the center of the genome. *J. Mol. Biol.* **241**: 651–662.
  113. Nowotny, M. 2009. Retroviral integrase superfamily: the structural perspective. *EMBO Rep.* **10**: 144–151.
  114. Charpentier, E. & J.A. Doudna. 2013. Biotechnology: rewriting a genome. *Nature* **495**: 50–51.
  115. Ohtani, N., H. Yanagawa, M. Tomita, *et al.* 2004. Cleavage of double-stranded RNA by RNase HI from a thermoacidophilic archaeon, *Sulfolobus tokodaii* 7. *Nucleic Acids Res.* **32**: 5809–5819.
  116. Simon, D.M. & S. Zimmerly. 2008. A diversity of uncharacterized reverse transcriptases in bacteria. *Nucleic Acids Res.* **36**: 7219–7229.
  117. Muotri, A.R., V.T. Chu, M.C. Marchetto, *et al.* 2005. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**: 903–910.
  118. Erwin, J.A., M.C. Marchetto & F.H. Gage. 2014. Mobile DNA elements in the generation of diversity and complexity in the brain. *Nat. Rev. Neurosci.* **15**: 497–506.
  119. Guo, H., D. Arambula, P. Ghosh & J.F. Miller. 2014. Diversity-generating retroelements in phage and bacterial genomes. *Microbiol. Spectr.* **2**. <https://doi.org/10.1128/microbiolspec.MDNA3-0029-2014>.
  120. Maksakova, I.A., D.L. Mager & D. Reiss. 2008. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cell. Mol. Life Sci.* **65**: 3329–3347.
  121. Dolinoy, D.C., D. Huang & R.L. Jirtle. 2007. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proc. Natl. Acad. Sci. USA* **104**: 13056–13061.
  122. Cropley, J.E., C.M. Suter, K.B. Beckmann, *et al.* 2006. Germ-line epigenetic modification of the murine A<sub>vy</sub> allele by nutritional supplementation. *Proc. Natl. Acad. Sci. USA* **103**: 17308–17312.
  123. Rakyán, V.K., S. Chong, M.E. Champ, *et al.* 2003. Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proc. Natl. Acad. Sci. USA* **100**: 2538–2543.
  124. Kazachenka, A., T.M. Bertozzi, M.K. Sjöberg-Herrera, *et al.* 2018. Identification, characterization, and heritability of murine metastable epialleles: implications for non-genetic inheritance. *Cell* **175**: 1259–1271.e13.
  125. Burgess, D.J. 2019. Testing the generality of non-genetic inheritance. *Nat. Rev. Genet.* **20**: 3.
  126. Waterland, R.A., M. Travisano & K.G. Tahiliani. 2007. Diet-induced hypermethylation at agouti viable yellow is not inherited transgenerationally through the female. *FASEB J.* **21**: 3380–3385.
  127. Rassoulzadegan, M. & F. Cuzin. 2018. Nutrition meets heredity: a case of RNA-mediated transmission of acquired characters. *Environ. Epigenet.* **4**: dv0006.

128. Ashe, A., A. Sapetschnig, E.M. Weick, *et al.* 2012. piRNAs can trigger a multigenerational epigenetic memory in the germ line of *C. elegans*. *Cell* **150**: 88–99.
129. Chen, L., J.Y. Chen, X. Zhang, *et al.* 2017. R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. *Mol. Cell* **68**: 745–757.e5.
130. Sanz, L.A., S.R. Hartono, Y.W. Lim, *et al.* 2016. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell* **63**: 167–178.
131. Brink, R.A., E.D. Styles & J.D. Axtell. 1968. Paramutation: directed genetic change. Paramutation occurs in somatic cells and heritably alters the functional state of a locus. *Science* **159**: 161–170.
132. Rassoulzadegan, M., V. Grandjean, P. Gounon, *et al.* 2006. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**: 469–474.
133. Wagner, K.D., N. Wagner, H. Ghanbarian, *et al.* 2008. RNA induction and inheritance of epigenetic cardiac hypertrophy in the mouse. *Dev. Cell* **14**: 962–969.
134. Grandjean, V., P. Gounon, N. Wagner, *et al.* 2009. The miR-124-Sox9 paramutation: RNA-mediated epigenetic control of embryonic and adult growth. *Development* **136**: 3647–3655.
135. Murashov, A.K., E.S. Pak, M. Koury, *et al.* 2016. Paternal long-term exercise programs offspring for low energy expenditure and increased risk for obesity in mice. *FASEB J.* **30**: 775–784.
136. Gapp, K., A. Jawaid, P. Sarkies, *et al.* 2014. Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nat. Neurosci.* **17**: 667–669.
137. Rodgers, A.B., C.P. Morgan, N.A. Leu, *et al.* 2015. Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. *Proc. Natl. Acad. Sci. USA* **112**: 13699–13704.
138. Carrapiço, F. 2010. How symbiogenic is evolution? *Theory Biosci.* **129**: 135–139.
139. Roossinck, M.J. 2011. The good viruses: viral mutualistic symbioses. *Nat. Rev. Microbiol.* **9**: 99–108.
140. Villarreal, L.P. 2016. Persistent virus and addiction modules: an engine of symbiosis. *Curr. Opin. Microbiol.* **31**: 70–79.
141. Margulis, L. 1970. *Origin of Eukaryotic Cells*. New Haven, CT: Yale University Press.
142. Margulis, L. 1993. *Symbiosis in Cell Evolution*. 2nd ed. New York, NY: W.H. Freeman and Co.
143. Emelyanov, V.V. 2001. Evolutionary relationship of Rickettsiae and mitochondria. *FEBS Lett.* **501**: 11–18.
144. Raven, J.A. & J.F. Allen. 2003. Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biol.* **4**: 209.
145. Blanda, V., A. Torina, F. La Russa, *et al.* 2017. A retrospective study of the characterization of *Rickettsia* species in ticks collected from humans. *Ticks Tick Borne Dis.* **8**: 610–614.
146. Kent, B.N., L. Salichos, J.G. Gibbons, *et al.* 2011. Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*) determined by targeted genome capture. *Genome Biol. Evol.* **3**: 209–218.
147. Wolf, Y.I. & E.V. Koonin. 2013. Genome reduction as the dominant mode of evolution. *Bioessays* **35**: 829–837.
148. Longdon, B. & F.M. Jiggins. 2012. Vertically transmitted viral endosymbionts of insects: do sigma viruses walk alone? *Proc. Biol. Sci.* **279**: 3889–3898.
149. Nakanishi, K., M. Hoshino, M. Nakai & Y. Kunimi. 2008. Novel RNA sequences associated with late male killing in *Homona magnanima*. *Proc. R. Soc. B* **275**: 1249–1254.
150. Spiegelman, S., I. Haruna, I.B. Holland, *et al.* 1965. The synthesis of a self-propagating and infectious nucleic acid with a purified enzyme. *Proc. Natl. Acad. Sci. USA* **54**: 919–927.
151. Kacian, D.L., D.R. Mills, F.R. Kramer, *et al.* 1972. A replicating RNA molecule suitable for a detailed analysis of extracellular evolution and replication. *Proc. Natl. Acad. Sci. USA* **69**: 3038–3042.
152. Suttle, C.A. 2005. Viruses in the sea. *Nature* **437**: 356–361.
153. Terry, S.N., L. Manganaro, A. Cuesta-Dominguez, *et al.* 2017. Expression of HERV-K108 envelope interferes with HIV-1 production. *Virology* **509**: 52–59.